# INVESTIGATION OF GENETIC VARIABILITY WITHIN AND BETWEEN LAKE WASHINGTON SOCKEYE SALMON POPULATIONS USING MICROSATELLITE MARKERS

**DRAFT Report submitted to City of Seattle by**

**Paul Bentzen and Ingrid Spies**

**University of Washington**
**Marine Molecular Biotechnology Laboratory**
**& School of Fisheries**
**3707 Brooklyn Avenue NE**
**Seattle, WA 98105-6715**

**January 2000**

e-mail: pbentzen@fish.washington.edu
Fax: 206-685-6651
Tel: 206-685-9994

## INTRODUCTION

We conducted genetic analyses of sockeye salmon and kokanee (*Oncorhynchus nerka*) from the Lake Washington drainage and two other Washington State drainages. Our goals were to initiate analyses of the genetic structure of L. Washington drainage *O. nerka* populations and the relationships of these populations to other sockeye and kokanee populations within the Pacific Northwest. A particular focus of the first phase of this investigation was an examination of genetic variation among samples of spawning sockeye collected at six intervals over the 1997 spawning season. The objective of this analysis was to determine whether Cedar R. sockeye show evidence of genetic differentiation over their protracted spawning season. We also examined genetic variation in Bear Creek sockeye, Issaquah Creek kokanee, Baker Lake sockeye, and Lake Whatcom kokanee as a first step towards describing the relationships of these and other Pacific Northwest populations. Hendry et al. (1996) previously concluded on the basis of limited allozyme data that Cedar R. sockeye are descended primarily from transplants of Baker L. sockeye, whereas Bear Creek sockeye are primarily of native origin. We re-evaluated these conclusions using DNA markers, and also evaluated the possibility that L. Whatcom kokanee may have contributed to current sockeye populations in the L. Washington drainage.

The genetic markers we examined in our analyses are microsatellites, DNA sequences comprised of 2-4 base pair (bp) motifs repeated in tandem arrays (e.g., CACACACA...) embedded in complex, non-repetitive DNA sequences. Microsatellites display a number of characteristics that make them extremely useful genetic markers: (1) microsatellites are abundant and dispersed on the chromosomes of all eukaryotic organisms. (2) They have high mutation rates (typically $10^{-3}$-$10^{-4}$) and consequently, high levels of allelic variability. (3) They are inherited as codominant Mendelian markers. (4) The genotypes of individual microsatellite arrays can readily be analyzed by amplifying them from genomic DNA samples via the polymerase chain reaction (PCR) and oligonucleotide primers complementary to the non-repetitive sequences immediately flanking the array. The sizes (and hence the number of repeat units) of the amplified microsatellite alleles are determined using electrophoresis on high-resolution polyacrylamide gels. (5) The DNA needed for analyses can be obtained from small tissue samples (e.g., fin clips or muscle) stored frozen, or at ambient temperatures in alcohol (EtOH). For all of these reasons, microsatellites have become the genetic markers of choice in a wide range of population and evolutionary genetic analyses of salmonid fishes and other organisms (reviewed in Wright and Bentzen 1994; O'Connell and Wright 1997).

## METHODS

Sockeye and kokanee were collected as indicated in Table 1. DNA was isolated from fin clips preserved in EtOH using a standard phenol-chloroform method (Sambrook et al. 1989). Six microsatellite loci, One1, One2, One8, One11, and One14 (Scribner et al. 1996) and Ots103 (Beacham et al. 1998) were amplified from DNA samples using two multiplex reactions as described in Woody (1998). Microsatellite alleles were resolved using an Applied Biosystems Inc. (ABI) 373A-XL automated DNA sequencer operated

in genescan mode, and microsatellite genotypes were scored with the help of ABI Genescan and Genotyper software as described in Olsen et al. 1996.

Observed and expected heterozygosity ($H_O$ and $H_E$) and FIS were calculated using Genepop 3.1d (Raymond and Rousset 1995). The significance of deviations of observed genotype frequencies from those expected under random mating (Hardy Weinberg Equilibrium, HWE) and heterogeneity in allele frequencies among samples were calculated using probability tests implemented in Genepop (Raymond and Rousset 1995a,b; Rousset and Raymond 1995). FST was calculated according to Weir and Cockerham (1984) and the significance of FST estimates was determined using a permutation test in which alleles were shuffled among samples, as implemented in Genetix v.4 (K. Belkhir; http://www.univ-montp2.fr/~genetix/constr.htm#download). The genetic distance measure, DLR (Paetkau et al. 1997) was computed using the calculator available at http://www.biology.ualberta.ca/jbrzusto/Doh.html. Assignment tests were carried out using the Bayesian approach of Rannala and Mountain (1997) as implemented in GeneClass (Cornuet et al. 1999). A dendrogram interpreting the genetic relationships of sockeye and kokanee samples was generated according to Felsenstein (1981) using PHYLIP (Felsenstein 1993). The resulting tree was drawn with the help of TreeView (Page 1996).

## RESULTS AND DISCUSSION

Genotyping success and genetic variability within population samples

We attempted to analyze genotypic variation at six microsatellite loci in 577 *O. nerka* from the Cedar River, Bear Creek, Baker Lake, Issaquah Creek and Lake Whatcom, and succeeded in obtaining genotypes for all or most loci in 546 samples. The remaining 31 samples yielded little or no genotypic data despite repeated efforts to extract DNA from these samples and to amplify the microsatellite loci. The majority of these genotyping failures occurred with samples collected from the Cedar R. on 12/03/97 and 12/22/97, and reflected poor preservation of the samples. In the case of the12/22/97 samples the poor preservation was apparently due to the fact that many of the samples were collected from spawned-out dead salmon in the river; these samples had likely undergone some decomposition prior to sampling and preservation of the fin clips in alcohol. The cause of the poor preservation of the 12/03/97 samples is unclear.

Analysis of the microsatellite loci revealed varying levels of genetic variability across the six loci and 13 population samples (Table 2, App. 1). The number of alleles per locus per population ranged from two (One1 in a number of samples) to 17 (Ots103 in Cedar 11/20/97 and Lake Whatcom kokanee) and averaged 7.9 across all loci and population samples. Expected heterozygosity (He) was similarly variable; it ranged from 0.06 (One1 in Whatcom kokanee) to 0.93 (Ots103 in Whatcom kokanee) and averaged 0.65 across all loci and populations.

We computed FIS and used probability tests (Rousset and Raymond 1995) to determine whether it differed significantly from zero for each locus in each population sample as a 'quality check' on the genotypic data. FIS measures the deviation of genotype frequencies from those expected from random mating. In addition to non-random mating, non-zero FIS values could arise from natural selection acting on genotypes, or from

genotyping artifacts derived from problems occurring in PCR amplification or electrophoretic analysis of microsatellite alleles. In the absence of additional lines of evidence, it is not possible to definitively determine which factors result in particular non-zero FIS values. However, non-random mating usually results in non-zero FIS values at several loci within a population sample, whereas genotyping artifacts and selection tend to be associated with particular loci and often occur in multiple population samples. Effects of natural selection on microsatellite genotypes are expected to be rare, because most microsatellites are not associated with phenotypic traits.

None of the population samples showed FIS values significantly different from zero at multiple loci; however, two loci did exhibit FIS values that were significantly different from zero in multiple population samples (Table 2). For one locus, One11, FIS values were negative (indicative of an excess of heterozygotes) in 12 of 13 population samples; in three samples (Cedar 10/06/97and 11/20/97, and Baker) these negative FIS values were significant ($P \leq 0.0017$). For this reason, and because One11 tended to amplify poorly in Cedar R. samples, One11 data were excluded from further analyses. Another locus, One2, exhibited significant positive FIS values (indicative of too few heterozygotes) in three samples (Bear Creek 1992, 1993 and Whatcom kokanee; $P \leq 0.0003$). This result suggests the possibility of a null (non-amplifying) One2 allele in these populations, and consequently One2 data were treated as missing data in some analyses involving these populations. Finally, one additional locus, One8, exhibited a large, statistically significant FIS value (0.38, $P \approx 0$) in a single sample (Cedar 12/22/97). This result may reflect a genotyping artifact that arose because of the poor quality of this sample, and as a conservative measure, One8 data were treated as 'missing' for this sample.

Genetic differentiation of 'temporal' population segments within the Cedar River

We conducted three series of tests to evaluate the null hypothesis that sockeye spawning in the Cedar R. (either naturally, or in the hatchery) comprise a single genetically homogenous assemblage. In the first series we used exact tests (Raymond and Rousset 1995) to test for heterogeneity of allele frequencies among samples collected from the Cedar R. at six intervals over the spawning season (Oct. 6-Dec. 22) in 1997. Heterogeneity in allele frequencies among all 1997 samples was significant at two loci (One2 and Ots103; $P \approx 0$ and $P = 0.0028$, respectively) following correction for five simultaneous tests, and was highly significant across all loci in a combined test (Fisher's method). The results were similar when we included Cedar R. samples collected in 1993 and 1995 in the analysis, except that in addition to One2 and Ots103, heterogeneity in allele frequencies was also significant for One1 ($P = 0.0123$). Tests of allele frequency heterogeneity among pairs of Cedar R. samples revealed that most pairs of samples differed significantly at one or two loci (Table 3, App. 2).

In the second series of tests we computed FST (a measure of the proportion of total genetic variance associated with population subdivision) for each pair of Cedar R. samples. FST can range from 0 (no population subdivision) to 1 (complete fixation of alternate alleles, indicating complete isolation of population segments). The mean pairwise FST for all comparisons among Cedar R. samples (including the 1993 and 1995 samples) was 0.009 (Table 4). We used permutation tests based on random shuffling of

alleles among population samples to test whether the estimated FST values were significantly greater than zero. Among 28 estimates of multilocus FST among pairs of Cedar R. samples, 21 were significantly greater than zero at $P \leq 0.05$ (Table 4).

In the third series of tests we tested our ability to assign individuals to sample dates based on their multilocus genotypes. We used a Bayesian approach suggested by Rannala and Mountain (1997) as implemented in the program GeneClass (Cornuet et al. 1999). In this method, individuals are assigned to the population in which their genotype has the highest probability. To avoid upward bias in assignment accuracy, we used the "take one out" option in GeneClass, in which the individual being assigned is removed from the database for its own population. Assignment accuracy varied among sampling dates (Fig. 1). Modal assignment dates for sockeye sampled 6-Oct, 20-Oct and 22-Dec were the correct dates. For the remaining sampling dates, the modal assignment date was either a tie between two consecutive dates in which one was the correct date (3-Nov) or else was a sampling date one time interval removed from the correct date (20-Nov and 3-Dec).

Overall, the results of the assignment tests suggested a tendency for sockeye to be assigned to dates on or close to the 'correct' date. To further evaluate this possibility, we compared the frequency distribution of 'assignment deviations' (number of individuals assigned for a given number of sampling intervals from the correct date) to the 'null' distribution of assignment deviations that would be expected if sockeye were assigned to sampling dates at random (Fig. 2). This comparison showed that more sockeye were assigned to the correct date than expected by chance (80 vs. 45), and fewer sockeye were assigned to dates two or more sampling intervals apart from the correct date than expected by chance. The difference between the two frequency distributions was significant ($\chi^2 = 45.26$, 5 df, $P < 0.001$).

Each of the three series of tests described above support the conclusion that Cedar R. sockeye spawned on different dates in 1997 did not form a genetically homogenous assemblage. We next tested the possibility that Cedar R. sockeye are structured in time in a manner analogous to the way many other populations are structured in space. A pattern of population structure exhibited by many populations that either occupy environments that are more or less continuous, or else habitats that are linked by high levels of gene flow, is one in which genetic differentiation increases with the degree of spatial separation. Such population structure is commonly referred to as isolation by distance (see Olsen et al. 1998 for an example involving pink salmon). We hypothesized that Cedar R. sockeye, with its protracted spawning period, is composed of multiple segments partially isolated by time rather than by distance. We used a Mantel test to test for a positive relationship between genetic distance and the amount of time separating samples of spawning sockeye. We used two measures of genetic distance, DLR (Paetkau et al. 1997), and an FST-based measure (Rousset 1997). Isolation by time was marginally significant for all 1997 Cedar R. samples using DLR ($P = 0.0531$) and $F_{ST}$ ($P = 0.0627$). We next tested whether the contribution of the two sexes to the weak isolation by time relationship was similar. Isolation by time was not significant for males using either genetic distance measure, but was significant for females (Figure 3; DLR, $P = 0.0047$; $F_{ST}$, $P = 0.0087$).

These results provide preliminary evidence that Cedar R. sockeye are structured genetically according to their spawning dates, such that sockeye spawning on similar

dates tend to be more similar genetically than they are to sockeye spawning substantially earlier or later in the season. It is unclear why female sockeye should display a significant isolation by time genetic structure and males should not. However, small sample sizes (which became even smaller when temporal samples were split according to sex) and the small number of loci assayed likely contributed considerable sampling error to the analyses. It is also possible that males really do exhibit weaker genetic differentiation by time, perhaps because they are able to spawn repeatedly over a longer time interval than females. Whether or not the pattern of genetic differentiation by time is similar for females and males, the differentiation may be maintained by natural selection. Cedar R. sockeye may be adapted to spawning at different times, with the consequence that those that spawn at the 'wrong' time (for their genetic makeup) experience reduced reproductive success. Evidence in support of this scenario comes from the fact that spawning date is known to be heritable in Pacific salmon (Siitonen and Gall 1989, McGregor et al. 1998, Sakamoto et al. 1999).

Genetic relationships of sockeye and kokanee populations

A long term goal of this study is to determine the relationships of sockeye and kokanee populations within the Lake Washington drainage, and the relationships of these populations to other populations of *O. nerka* in Washington and the Pacific Northwest. Full realization of this goal will require data from many more samples and genetic markers than were available for this phase of the study. For the present report we confine our genetic comparisons to the following populations (where S and K denote sockeye and kokanee, respectively): Cedar R. (S), Bear Creek (S), Baker L. (S), L. Whatcom (K) and Issaquah Creek (K).

Tests of allele frequency heterogeneity (Table 3) and $F_{ST}$ (Table 4) revealed significant genetic differentiation among all populations. Within the Lake Washington drainage, the magnitude of genetic differentiation followed a pattern that might be expected based on location and phenotype. Least differentiated were the various samples obtained from the Cedar R. on various dates (as noted earlier, mean $F_{ST}$ = 0.009). Next in order of differentiation was Bear (S) vs. Cedar (S) (mean $F_{ST}$ = 0.018), then Issaquah (K) vs. Bear (S) (mean $F_{ST}$ = 0.078) and finally Issaquah (K) vs. Cedar (S) (mean $F_{ST}$ = 0.102).

Mean $F_{ST}$ between the Baker (S) sample and Cedar (S), Bear (S) and Issaquah (K) samples was 0.067, 0.095 and 0.215, respectively. Finally, $F_{ST}$ between Baker (S) and Whatcom (K) was 0.156.

Comparisons of Whatcom (K) to Lake Washington populations produced some surprising results. Mean $F_{ST}$ for Whatcom (K) vs. Issaquah (K), Bear (S) and Cedar (S) was 0.074, 0.019 and 0.054, respectively. The close genetic similarity between Bear (S) and Whatcom (K) suggests the possibility of introgression of Whatcom kokanee genes into L. Washington drainage *O. nerka* populations following stocking of Lake Whatcom kokanee in the L. Washington system early in the 20$^{th}$ century. If this interpretation is correct, the relatively large genetic distance between the Whatcom (K) and Issaquah (K) samples is unexpected. However, it could be an artifact of the small sample size (N= 13) of the Issaquah (K) sample.

We used a maximum likelihood (ML) approach (Felsenstein 1981) to interpret the genetic relationships of the poplation samples in a dendrogram (Figure 4). This model approximates the effects of genetic drift and gene flow (straying/migration). The ML tree closely follows the pattern of FST estimates described above. Within the L. Washington drainage, the most genetically divergent sample is Issaquah (K). As noted for the FST results, the long branch length leading to Issaquah (K) may be an artifact of small sample size for this population. Also apparent in the ML tree is the genetic similarity of Bear (S) and Whatcom (K). Finally, the Cedar (S) samples occupy a central cluster between the Baker (S) and other L. Washington samples, suggesting the possibility of mixed origins for Cedar R. sockeye involving both Baker L. sockeye and ancestors common to other L. Washington drainage populations.

Our results can be compared to those of Hendry et al. (1996). They surveyed genetic variation at 13 allozyme loci (four of which were variable enough to provide useful data) in several populations with the goal of identifying the origins of L. Washington drainage sockeye. Many of the samples they studied, Baker L., Cedar R. 1992 and 1993, Bear Creek 1992 and 1993, and Issaquah kokanee, were included in our study as well. They found evidence that L. Washington drainage sockeye populations belong to two groups: 'Group 1' comprised Cedar R., L. Washington beach and Issaquah Creek sockeye and 'Group 2' was made up of Bear and Cottage Creek fish. Since Group 1 was more closely related to Baker L. sockeye than to Group 2, they concluded that these populations were derived primarily from introductions of Baker L. sockeye. Group 2, in turn, was more similar to a kokanee sample (comprised mainly of Issaquah fish with three additional kokanee from Bear and Cottage Creeks) than it was to Group 1. On the basis of the assumption that the kokanee were representative of native L. Washington drainage *O. nerka*, they concluded that Group 2 was primarily of native origin. Our microsatellite data are only partially consistent with Hendry et al.'s results, and suggest rather different conclusions. We also found evidence that Cedar R. sockeye were related to Baker sockeye; however, in our analysis Cedar R. sockeye were much more closely related to Bear Creek sockeye than to the Baker Lake population. Further, we found evidence that L. Whatcom kokanee (not studied by Hendry et al.) may be related to L. Washington drainage sockeye populations. Estimates of genetic differentiation between Bear Creek and L. Whatcom were similar to those between Bear Creek and Cedar R. (mean $F_{ST}$ = 0.019 and 0.018, respectively). Even the Cedar R. population was no more similar to Baker L. sockeye than to L. Whatcom kokanee (mean $F_{ST}$ = 0.067 and 0.054, respectively). Hence, our results suggest the possibility that the contribution of L. Whatcom kokanee to L. Washington drainage sockeye populations may have been at least as great as that of Baker L. sockeye. They also raise some doubts about Hendry et al.'s conclusion that Group 2 sockeye are primarily of native origin.

## SUMMARY AND SUGGESTIONS FOR FURTHER WORK

Our microsatellite analyses led to two important results. First, we found evidence that sockeye spawning in the Cedar R. do not represent a single, genetically homogenous assemblage. Instead, there is evidence that the population is genetically structured by time of spawning, such that the genetic differentiation of fish spawning on different dates is positively correlated with the amount of time separating them. Our second important

observation was the remarkable genetic similarity of L. Whatcom kokanee to Bear Creek sockeye, and to a lesser but still notable extent, Cedar R. sockeye.

Both results should be regarded as provisional, and could be substantially bolstered (or conceivably, refuted) by further work. The genetic isolation by time effect seemed to be stronger for females than for males; however, it is unclear whether this difference between the sexes was real or an artifact of sampling. In particular, our analyses suffered from small sample sizes in critical end-of-season samples (12/03/97 and 12/22/97) brought about by poor preservation of the tissue. It would be desirable to repeat these analyses with new samples. With this goal in mind, we collected 532 new samples of spawning sockeye from the hatchery in fall 1999. We collected 100 fish per week (50 of each sex) when possible to improve the resolution of future analyses conducted on these samples. Unfortunately, the small size of the 1999 Cedar R. sockeye run made it impossible to meet our sample objectives after the beginning of November. It would also be of interest to examine early and late spawning sockeye from the Issaquah Creek population.

Our analysis of the genetic relationships of L. Washington drainage sockeye and kokanee populations would also benefit from inclusion of further samples. Clearly, it would be desirable to examine Issaquah Creek sockeye, and additional samples of Bear and Cottage Creek sockeye, as well as whatever kokanee samples can be obtained from the drainage. It would also be useful to include appropriate 'outgroup' populations of sockeye and kokanee in the analysis.

Both the 'isolation by time' analysis of the Cedar R. population and the broader phylogenetic analysis of key sockeye and kokanee populations would also benefit from inclusion of more genetic loci. When we began these analyses, the number of microsatellites markers available for sockeye was limited; however, recently this situation has improved markedly. New sockeye microsatellite loci developed by the Alaska Dept. of Fish and Game and numerous other new microsatellites developed in the Bentzen laboratory have created a wide assortment of new markers for this species (J. Olsen, pers. com., and Bentzen, unpub. data). The addition of new markers is critical, because the precision of genetic differentiation estimates, and the statistical support for particular branching orders in phylogenetic trees are both even more dependent on the number of loci examined than the size of the population samples. This was a key weakness in the study by Hendry et al. (1996) as well, since they were only able to examine four informative loci--too few to robustly support the conclusions that they reached. Moreover, their study relied critically on alleles that could not be directly detected (only inferred statistically), and was also weakened by the possibility that their markers were influenced by natural selection.

It would also be possible to examine mitochondrial DNA (mtDNA). By virtue of its maternal inheritance, inclusion of mtDNA might be particularly informative in terms of resolving the question of the origins of the various L. Washington sockeye and kokanee populations. If it is assumed that hybridization between sockeye and kokanee is most likely to involve female sockeye mating with male kokanee, and it is further hypothesized that both Baker L. sockeye and L. Whatcom kokanee contributed to existing populations in the drainage, one might predict that Baker L. mtDNA genotype(s) should predominate in Cedar R. (and Bear Creek?) sockeye. On the other hand, similar logic would suggest the possibility that L. Whatcom mtDNA genotypes might occur in L.

Washington kokanee. Finally, it might also be possible to detect unique L. Washington mtDNA genotypes (i.e., genotypes that are nonexistent or rare in the Baker L. and L. Whatcom) that would offer support for a 'native' component to extant L. Washington drainage populations. At present there are no data supporting (or refuting) the existence of such informative mtDNA genotypes; however, it would not be difficult to look for them.

Our results generally agree with those of Hendry et al. (1996) inasmuch as they provide further evidence that populations of sockeye and kokanee in the L. Washington drainage are genetically differentiated from each other. This result has at least two important implications. First, it suggests the possibility that if further enhancement of the Cedar R. population were to lead to increased numbers of Cedar R. fish straying into other populations, those populations might be genetically altered by interbreeding with Cedar R. sockeye. Second, it might be possible to identify Cedar R. strays in other streams on the basis of their microsatellite genotypes, and even in the absence of hatchery marks.

## REFERENCES

Beacham TD, Margolis, L, and Nelson, RJ. 1998. A comparison of methods of stock identification for sockeye salmon, *Oncorhynchus nerka*, in Barkley Sound, British Columbia. *North Pacific Anadromous Fish Commission Bulletin* 1:227-239.

Cornuet JM, Piry S, Luikart G, Estoup A, Solignac, M. 1999. Comparison of methods employing multilocus genotypes to select or exclude populations as origins of individuals. (submitted article)

Felsenstein J. 1981. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. *Evolution*, 35:1229-1242.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. *Distributed by the author. Department of Genetics, University of Washington, Seattle.*

Hendry, AP, Quinn, TP, and Utter, FM. 1996. Genetic evidence for the persistence and divergence of native and introduced sockeye salmon (*Oncorhynchus nerka*) with Lake Washington, Washington. *Can. J. Fish. Aquat. Sci.* 53:823-832.

McGregor, AJ, Lane, S, Thomason, MA, Zhivotovsky, LA, Smoker, WW, Gharrett, AJ. 1998. *NPAFC Bulletin* 1:262-273.

O'Connell M and Wright JM. 1997. Microsatellite DNA in fishes. *Reviews in Fish Biology and Fisheries* 7:331-363.

Olsen JB, Wenburg JK, and Bentzen P. 1996. Semiautomated multilocus genotypic of Pacific salmon (*Oncorhynchus spp.*) using microsatellites. *Molecular Marine Biology and Biotechnology* 5:259-272.

Paetkau D, Waits LP, Clarkson PL, Craighead L, and Strobeck C. 1997. An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics* 147:1943-1957.

Page, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**: 357-358.

Olsen, JB, Seeb, LW, Bentzen, P, and Seeb, JE. 1998. Genetic interpretation of broad-scale microsatellite polymorphism in odd-year pink salmon. *Trans. Am. Fish. Soc.* 127:535-550.

Rannala, B, and Mountain, JL. 1997. Detecting immigration by using multilocus genotypes. *Proc. Natl. Acad. Sci. USA* 94: 9197-9221.

Raymond M and Rousset F. 1995a. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86:248-249.

Raymond M and Rousset F. 1995b. An exact test for population differentiation. *Evolution* 49:1280-1283.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145:1219-1228.

Rousset, F, and Raymond, M. 1995. Testing heterozygote excess and deficiency. *Genetics* 140:1413-1419.

Sakamoto,T, Danzmann, RG, Okamoto, N, Ferguson, MM, and Ihssen, PE. 1999. Linkage analysis of quantitative trait loci associated with spawning time in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 173:33-43.

Sambrook, J, Fritsch, EF, and Maniatis, T. 1989. Molecular Cloning: A laboratory manual (2nd ed.) Cold Spring Harbor Laboratory Press.

Scribner KT, Gust JR, and Fields RL. 1996. Isolation and characterization of novel salmon microsatellite loci: cross-species amplification and population genetic applications. *Canadian Journal of Fisheries and Aquatic Sciences* 53:833-841.

Siitonen, L, and Gall, GAE. 1989. Response t selection for early spawn date in rainbow trout, *Salmo gairdneri. Aquaculture* 78:153-161.

Weir BS and Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358-1370.

Woody, CA. 1998. Ecological, morphological, genetic and life history comparison of two sockeye salmon populations, Tustumena Lake, Alaska. Ph.D. thesis, School of Fisheries, University of Washington.

Wright JM and Bentzen P. 1994. Microsatellites: genetic markers for the future. *Reviews in Fish Biology and Fisheries* 4:384-388.

Table 1. Sockeye and kokanee samples used for genetic analyses.

| Population sample | Date collected | Type | N* |
|---|---|---|---|
| Cedar River | 10/6/97 | sockeye | 46 |
| Cedar River | 10/20/97 | sockeye | 51 |
| Cedar River | 11/3/97 | sockeye | 53 |
| Cedar River | 11/20/97 | sockeye | 50 |
| Cedar River | 12/3/97 | sockeye | 44 |
| Cedar River | 12/22/97 | sockeye | 45 |
| Cedar River | 11/19/93-12/3/93 | sockeye | 48 |
| Cedar River | 1995 | sockeye | 39 |
| Bear Creek | 10/5/92-10/21/92 | sockeye | 19 |
| Bear Creek | 9/28/93-11/5/93 | sockeye | 48 |
| Baker Lake | 1993 | sockeye | 42 |
| Lake Whatcom | 1999 | kokanee | 79 |
| Issaquah Creek | 1993 | kokanee | 13 |
| | | TOTAL | 577 |

* N is the number of samples from which DNA was extracted and genetic analysis attempted. The actual number of samples from which data was obtained varied. See Table 2 and App. 1 for further details.

Table 2. Summary statistics for microsatellite loci in sockeye and kokanee population samples. Abbreviations are as follows: N, number of fish genotyped at a locus; *A* , number of alleles; *Ho* , observed heterozygosity; *He* , expected heterozygosity; *P* , probability that Fis departs from zero (values shown in bold are significant following correction for multiple tests).

| Population sample | | One1 | One11 | One14 | One2 | One8 | Ots103 | MEAN | MEAN (excl. One11) |
|---|---|---|---|---|---|---|---|---|---|
| Cedar 10/6/97 | N | 45 | 27 | 44 | 44 | 43 | 43 | 41.0 | 43.8 |
| | *A* | 2 | 3 | 6 | 12 | 11 | 13 | 7.83 | 8.80 |
| | *Ho* | 0.53 | 0.85 | 0.59 | 0.84 | 0.86 | 0.88 | 0.76 | 0.74 |
| | *He* | 0.45 | 0.57 | 0.60 | 0.88 | 0.82 | 0.87 | 0.70 | 0.72 |
| | *Fis* | -0.19 | -0.50 | 0.02 | 0.05 | -0.06 | -0.02 | -0.12 | -0.04 |
| | *P* | 0.3127 | **0.0017** | 0.3604 | 0.2644 | 0.9788 | 0.8957 | | |
| Cedar 10/20/1997 | N | 51 | 31 | 51 | 51 | 51 | 51 | 47.7 | 51.0 |
| | *A* | 2 | 3 | 10 | 14 | 9 | 13 | 8.50 | 9.60 |
| | *Ho* | 0.45 | 0.68 | 0.61 | 0.75 | 0.92 | 0.88 | 0.71 | 0.72 |
| | *He* | 0.39 | 0.54 | 0.69 | 0.86 | 0.82 | 0.88 | 0.70 | 0.73 |
| | *Fis* | -0.15 | -0.26 | 0.12 | 0.13 | -0.13 | 0.00 | -0.05 | 0.00 |
| | *P* | 0.4695 | 0.0761 | 0.437 | 0.0611 | 0.2837 | 0.4368 | | |
| Cedar 11/03/97 | N | 50 | 23 | 50 | 50 | 50 | 49 | 45.3 | 49.8 |
| | *A* | 3 | 3 | 8 | 11 | 10 | 14 | 8.17 | 9.20 |
| | *Ho* | 0.48 | 0.65 | 0.82 | 0.82 | 0.88 | 0.84 | 0.75 | 0.77 |
| | *He* | 0.40 | 0.51 | 0.70 | 0.87 | 0.82 | 0.90 | 0.70 | 0.74 |
| | *Fis* | -0.19 | -0.28 | -0.18 | 0.06 | -0.08 | 0.07 | -0.10 | -0.06 |
| | *P* | 0.0546 | 0.2832 | 0.6964 | 0.238 | 0.5371 | 0.5796 | | |
| Cedar 11/20/97 | N | 49 | 38 | 49 | 46 | 46 | 46 | 45.7 | 47.2 |
| | *A* | 2 | 3 | 7 | 12 | 10 | 17 | 8.50 | 9.60 |
| | *Ho* | 0.31 | 0.82 | 0.78 | 0.89 | 0.78 | 0.89 | 0.74 | 0.73 |
| | *He* | 0.32 | 0.56 | 0.66 | 0.86 | 0.81 | 0.91 | 0.68 | 0.71 |
| | *Fis* | 0.03 | -0.48 | -0.19 | -0.03 | 0.03 | 0.02 | -0.10 | -0.03 |
| | *P* | 1 | **0.0005** | 0.7792 | 0.1477 | 0.137 | 0.1876 | | |
| Cedar 12/3/1997 | N | 35 | 25 | 34 | 33 | 31 | 29 | 31.2 | 32.4 |
| | *A* | 2 | 3 | 5 | 11 | 10 | 15 | 7.67 | 8.60 |
| | *Ho* | 0.43 | 0.64 | 0.74 | 0.88 | 0.71 | 0.93 | 0.72 | 0.74 |
| | *He* | 0.43 | 0.55 | 0.72 | 0.85 | 0.78 | 0.91 | 0.71 | 0.74 |
| | *Fis* | -0.01 | -0.18 | -0.02 | -0.04 | 0.10 | -0.02 | -0.03 | 0.00 |
| | *P* | 1 | 0.0364 | 0.09 | 0.92 | 0.11 | 0.2492 | | |
| Cedar 12/22/97 | N | 36 | 32 | 36 | 33 | 29 | 33 | 33.2 | 33.4 |
| | *A* | 3 | 3 | 6 | 15 | 6 | 14 | 7.83 | 8.80 |
| | *Ho* | 0.33 | 0.59 | 0.78 | 0.85 | 0.28 | 0.91 | 0.62 | 0.63 |
| | *He* | 0.33 | 0.57 | 0.64 | 0.90 | 0.44 | 0.90 | 0.63 | 0.64 |
| | *Fis* | -0.02 | -0.05 | -0.23 | 0.05 | 0.38 | -0.02 | 0.02 | 0.04 |
| | *P* | 1 | 0.5879 | 0.1133 | 0.1601 | **0** | 0.1689 | | |

Table 2. Continued.



| Population sample | | One1 | One11 | One14 | One2 | One8 | Ots103 | MEAN | MEAN (excl. One11) |
|---|---|---|---|---|---|---|---|---|---|
| Cedar 1993 | N | 46 | 30 | 47 | 47 | 46 | 44 | 43.3 | 46.0 |
| | *A* | 2 | 4 | 6 | 12 | 12 | 16 | 8.67 | 9.60 |
| | *Ho* | 0.26 | 0.73 | 0.68 | 0.81 | 0.91 | 0.89 | 0.71 | 0.71 |
| | *He* | 0.23 | 0.52 | 0.67 | 0.85 | 0.85 | 0.90 | 0.67 | 0.70 |
| | *Fis* | -0.14 | -0.41 | -0.02 | 0.05 | -0.08 | 0.02 | -0.10 | -0.03 |
| | *P* | 1 | 0.0404 | 0.6108 | 0.34 | 0.7407 | 0.6335 | | |
| Cedar 1995 | N | 36 | 27 | 36 | 34 | 36 | 36 | 34.2 | 35.6 |
| | *A* | 3 | 3 | 5 | 12 | 10 | 15 | 8.00 | 9.00 |
| | *Ho* | 0.53 | 0.48 | 0.61 | 0.88 | 0.83 | 0.86 | 0.70 | 0.74 |
| | *He* | 0.42 | 0.48 | 0.63 | 0.87 | 0.83 | 0.89 | 0.69 | 0.73 |
| | *Fis* | -0.26 | -0.01 | 0.03 | -0.02 | -0.01 | 0.04 | -0.04 | -0.04 |
| | *P* | 0.1466 | 1 | 0.4116 | 0.928 | 0.2039 | 0.2523 | | |
| Bear Crk. 1993 | N | 46 | 22 | 45 | 25 | 24 | 47 | 34.8 | 37.4 |
| | *A* | 3 | 4 | 11 | 11 | 9 | 15 | 8.83 | 9.80 |
| | *Ho* | 0.26 | 0.64 | 0.60 | 0.76 | 0.79 | 0.85 | 0.65 | 0.65 |
| | *He* | 0.34 | 0.50 | 0.62 | 0.90 | 0.72 | 0.90 | 0.66 | 0.70 |
| | *Fis* | 0.22 | -0.29 | 0.04 | 0.16 | -0.11 | 0.06 | 0.01 | 0.07 |
| | *P* | 0.067 | 0.5353 | 0.2493 | **0** | 0.6775 | 0.1825 | | |
| Bear Crk. 1992 | N | 17 | 14 | 16 | 11 | 14 | 17 | 14.8 | 15.0 |
| | *A* | 3 | 2 | 5 | 11 | 7 | 13 | 6.83 | 7.80 |
| | *Ho* | 0.18 | 0.36 | 0.50 | 0.55 | 0.79 | 0.94 | 0.55 | 0.59 |
| | *He* | 0.17 | 0.39 | 0.50 | 0.93 | 0.78 | 0.92 | 0.61 | 0.66 |
| | *Fis* | -0.04 | 0.09 | 0.00 | 0.42 | -0.01 | -0.03 | 0.07 | 0.07 |
| | *P* | 1 | 1 | 1 | **0.0003** | 0.9779 | 0.041 | | |
| Baker 1993 | N | 42 | 42 | 42 | 34 | 40 | 35 | 39.2 | 38.6 |
| | *A* | 2 | 3 | 8 | 11 | 6 | 13 | 7.17 | 8.00 |
| | *Ho* | 0.50 | 0.88 | 0.48 | 0.76 | 0.73 | 0.91 | 0.71 | 0.68 |
| | *He* | 0.50 | 0.61 | 0.46 | 0.80 | 0.72 | 0.87 | 0.66 | 0.67 |
| | *Fis* | 0.01 | -0.45 | -0.03 | 0.05 | -0.01 | -0.05 | -0.08 | -0.01 |
| | *P* | 1 | **0** | 0.1217 | 0.1586 | 0.0543 | 0.0421 | | |
| Whatcom kokanee | N | 78 | 77 | 77 | 74 | 76 | 74 | 76.0 | 75.8 |
| | *A* | 3 | 3 | 16 | 11 | 7 | 17 | 9.50 | 10.80 |
| | *Ho* | 0.06 | 0.26 | 0.56 | 0.51 | 0.67 | 0.86 | 0.49 | 0.53 |
| | *He* | 0.06 | 0.25 | 0.58 | 0.76 | 0.61 | 0.93 | 0.53 | 0.59 |
| | *Fis* | -0.02 | -0.03 | 0.03 | 0.33 | -0.10 | 0.07 | 0.04 | 0.06 |
| | *P* | 1 | 1 | 0.5131 | **0** | 0.142 | 0.4981 | | |
| Issaquah kokanee 1993 | N | 13 | 11 | 13 | 13 | 13 | 12 | 12.5 | 12.8 |
| | *A* | 2 | 3 | 7 | 8 | 3 | 8 | 5.17 | 5.60 |
| | *Ho* | 0.08 | 0.55 | 0.77 | 0.92 | 0.31 | 0.83 | 0.58 | 0.58 |
| | *He* | 0.08 | 0.44 | 0.71 | 0.82 | 0.34 | 0.78 | 0.53 | 0.54 |
| | *Fis* | 0.00 | -0.26 | -0.09 | -0.13 | 0.09 | -0.08 | -0.08 | -0.04 |
| | *P* | - | 1 | 0.6477 | 0.22 | 0.1652 | 0.8109 | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| **mean A** | 2.46 | 3.08 | 7.69 | 11.62 | 8.46 | 14.08 |
| **mean He** | 0.32 | 0.50 | 0.63 | 0.86 | 0.72 | 0.89 |
| **mean Fis** | -0.06 | -0.24 | -0.04 | 0.08 | 0.00 | 0.00 |

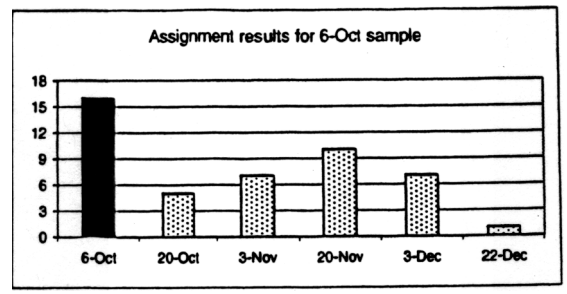
Table 3. Results of tests of allelic differentiation in pairwise comparisons of sockeye samples. Shown are the number of tests (out of 5) that were significant at P <= 0.001 (**bold**) or *P* <= 0.05 (*italic*). Detailed results of the pairwise tests are in App. 2. Cells filled in grey indicate comparisons within the Cedar River.

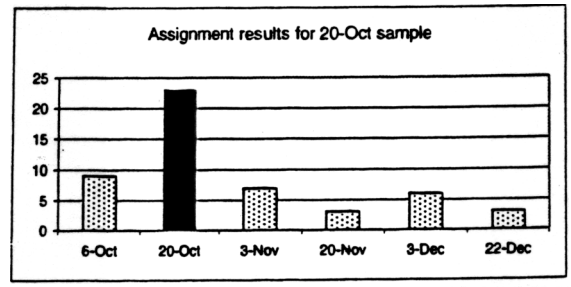
| | Whatcom kokanee | Issaquah kokanee | Cedar 12/22/1997 | Cedar 10/20/1997 | Cedar 10/6/1997 | Cedar 11/3/1997 | Cedar 11/20/1997 | Cedar 12/3/1997 | Cedar 1995 | Cedar 1993 | Baker 1993 | Bear 1993 | Bear 1992 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Whatcom kokanee | ---- | | | | | | | | | | | | |
| Issaquah kokanee | 4,*0* | ---- | | | | | | | | | | | |
| Cedar 12/22/1997 | 4,*0** | 3,*0** | ---- | | | | | | | | | | |
| Cedar 10/20/1997 | 5,*0* | 4,*1* | 0,*2** | ---- | | | | | | | | | |
| Cedar 10/6/1997 | 5,*0* | 4,*1* | 2,*1** | 1,*1* | ---- | | | | | | | | |
| Cedar 11/3/1997 | 5,*0* | 5,*0* | 1,*1** | 1,*0* | 0,*1* | ---- | | | | | | | |
| Cedar 11/20/1997 | 4,*1* | 3,*2* | 1,*1** | 1,*0* | 0,*2* | 0,*0* | ---- | | | | | | |
| Cedar 12/3/1997 | 4,*1* | 5,*0* | 1,*0** | 1,*0* | 0,*1* | 0,*0* | 0,*0* | ---- | | | | | |
| Cedar 1995 | 5,*0* | 3,*1* | 0,*2** | 1,*0* | 1,*1* | 1,*0* | 1,*0* | 1,*0* | ---- | | | | |
| Cedar 1993 | 4,*1* | 4,*1* | 0,*1** | 1,*1* | 1,*3* | 1,*1* | 0,*1* | 0,*2* | 0,*1* | ---- | | | |
| Baker 1993 | 5,*0* | 5,*0* | 4,*0** | 3,*2* | 2,*0* | 4,*1* | 3,*2* | 3,*2* | 3,*2* | 4,*1* | ---- | | |
| Bear 1993 | 4,*1* | 1,*4* | 1,*2** | 1,*3* | 1,*3* | 0,*4* | 0,*4* | 0,*4* | 0,*4* | 0,*3* | 3,*1* | ---- | |
| Bear 1992 | 2,*2* | 2,*1* | 1,*1** | 0,*3* | 2,*1* | 2,*1* | 1,*3* | 1,*1* | 1,*3* | 1,*2* | 3,*2* | 0,*0* | ---- |

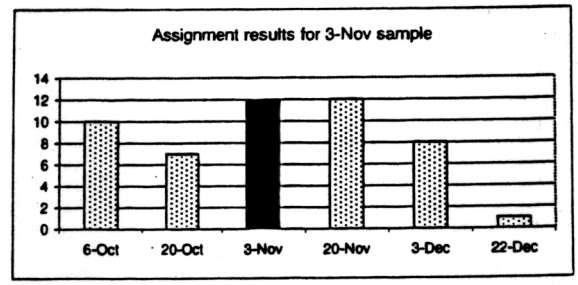
*Only 4 loci were tested for comparisons involving Cedar 12/22/97

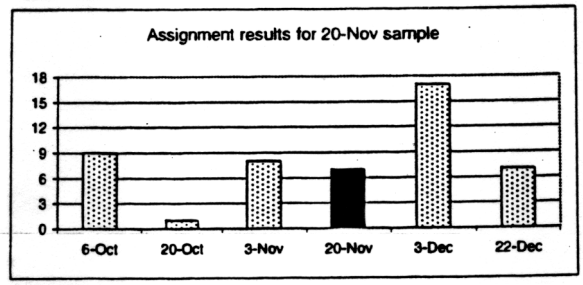
Table 4. Pairwise Fst estimates (above diagonal) and estimates of the probability that Fst is not greater than zero (below diagonal). Values shown in **bold** are significant at $P <= 0.001$, and values shown in *italics* are significant at $P <= 0.05$.

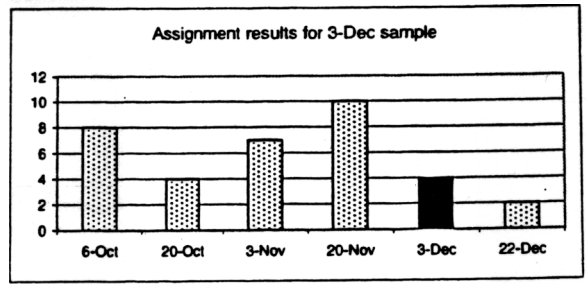
| | Whatcom kokanee | Issaquah kokanee | Cedar 12/22/1997 | Cedar 10/20/1997 | Cedar 10/6/1997 | Cedar 11/3/1997 | Cedar 11/20/1997 | Cedar 12/3/1997 | Cedar 1995 | Cedar 1993 | Baker 1993 | Bear 1993 | Bear 1992 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Whatcom kokanee | ---- | 0.074 | 0.040 | 0.057 | 0.071 | 0.056 | 0.040 | 0.059 | 0.061 | 0.051 | 0.156 | *0.017* | *0.021* |
| Issaquah kokanee | 0 | ---- | 0.079 | 0.106 | 0.127 | 0.102 | 0.094 | 0.102 | 0.097 | 0.110 | 0.215 | 0.062 | *0.094* |
| Cedar 12/22/1997 | 0 | 0 | ---- | *0.011* | 0.033 | *0.015* | *0.014* | *0.019* | *0.009* | 0.006 | 0.082 | 0.021 | 0.034 |
| Cedar 10/20/1997 | 0 | 0 | 0.017 | ---- | *0.007* | *0.008* | *0.011* | 0.007 | 0.004 | *0.008* | 0.057 | *0.016* | *0.019* |
| Cedar 10/6/1997 | 0 | 0 | 0 | 0.029 | ---- | *0.005* | *0.008* | 0.005 | *0.015* | 0.025 | 0.035 | 0.026 | 0.034 |
| Cedar 11/3/1997 | 0 | 0 | 0.002 | 0.012 | 0.053 | ---- | -0.001 | -0.003 | *0.007* | *0.013* | 0.064 | *0.015* | 0.033 |
| Cedar 11/20/1997 | 0 | 0 | 0.005 | 0.003 | 0.023 | 0.578 | ---- | 0.000 | *0.008* | *0.006* | 0.073 | 0.008 | *0.020* |
| Cedar 12/3/1997 | 0 | 0 | 0.005 | 0.076 | 0.102 | 0.774 | 0.512 | ---- | 0.007 | *0.012* | 0.069 | *0.013* | *0.029* |
| Cedar 1995 | 0 | 0 | 0.029 | 0.129 | 0.002 | 0.026 | 0.022 | 0.085 | ---- | *0.00704* | 0.069 | *0.015* | *0.029* |
| Cedar 1993 | 0 | 0 | 0.074 | 0.018 | 0 | 0.003 | 0.052 | 0.013 | 0.027 | ---- | 0.088 | 0.018 | *0.026* |
| Baker 1993 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ---- | 0.095 | 0.095 |
| Bear 1993 | 0.002 | 0 | 0.001 | 0.003 | 0 | 0.002 | 0.056 | 0.027 | 0.004 | 0 | 0 | ---- | 0.005 |
| Bear 1992 | 0.007 | 0.002 | 0 | 0.02 | 0.001 | 0 | 0.007 | 0.006 | 0.003 | 0.004 | 0 | 0.239 | ---- |



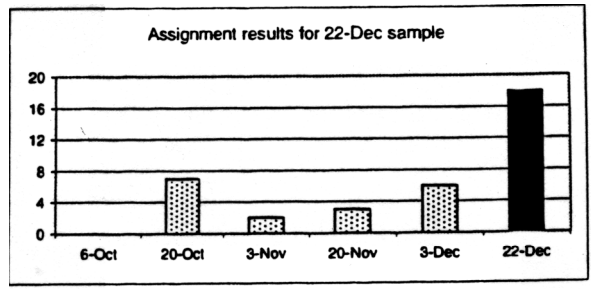




Figure 1. Results of assignment tests for 1997 Cedar R. sockeye samples. Histograms show the number of fish assigned to each sample date.

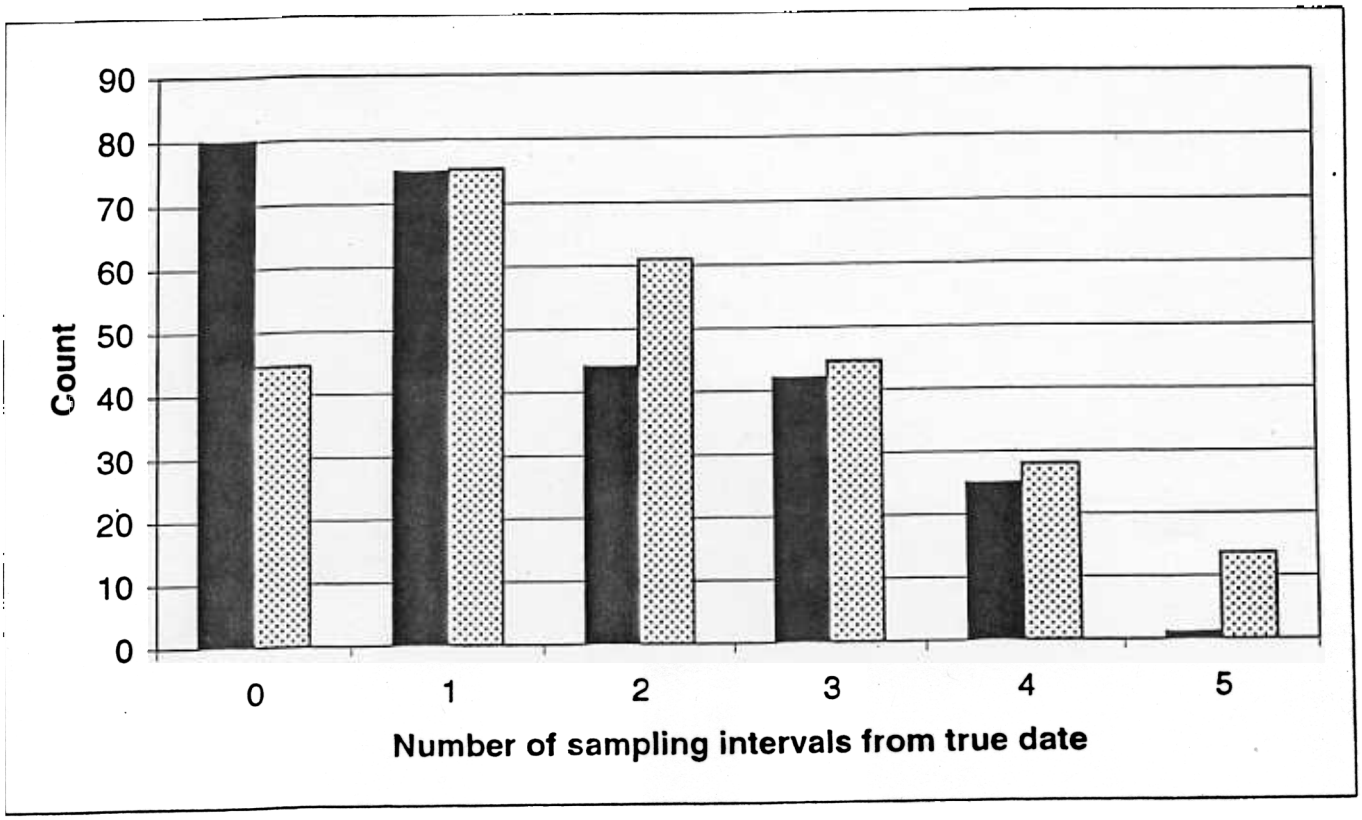



Figure 2. Frequency of assignment of sockeye salmon to sampling dates vs. the number of sampling intervals from the true date for each individual. Black bars show the observed distribution and the stippled bars show the expected null distribution of assuming random assignment of individuals to sampling date.

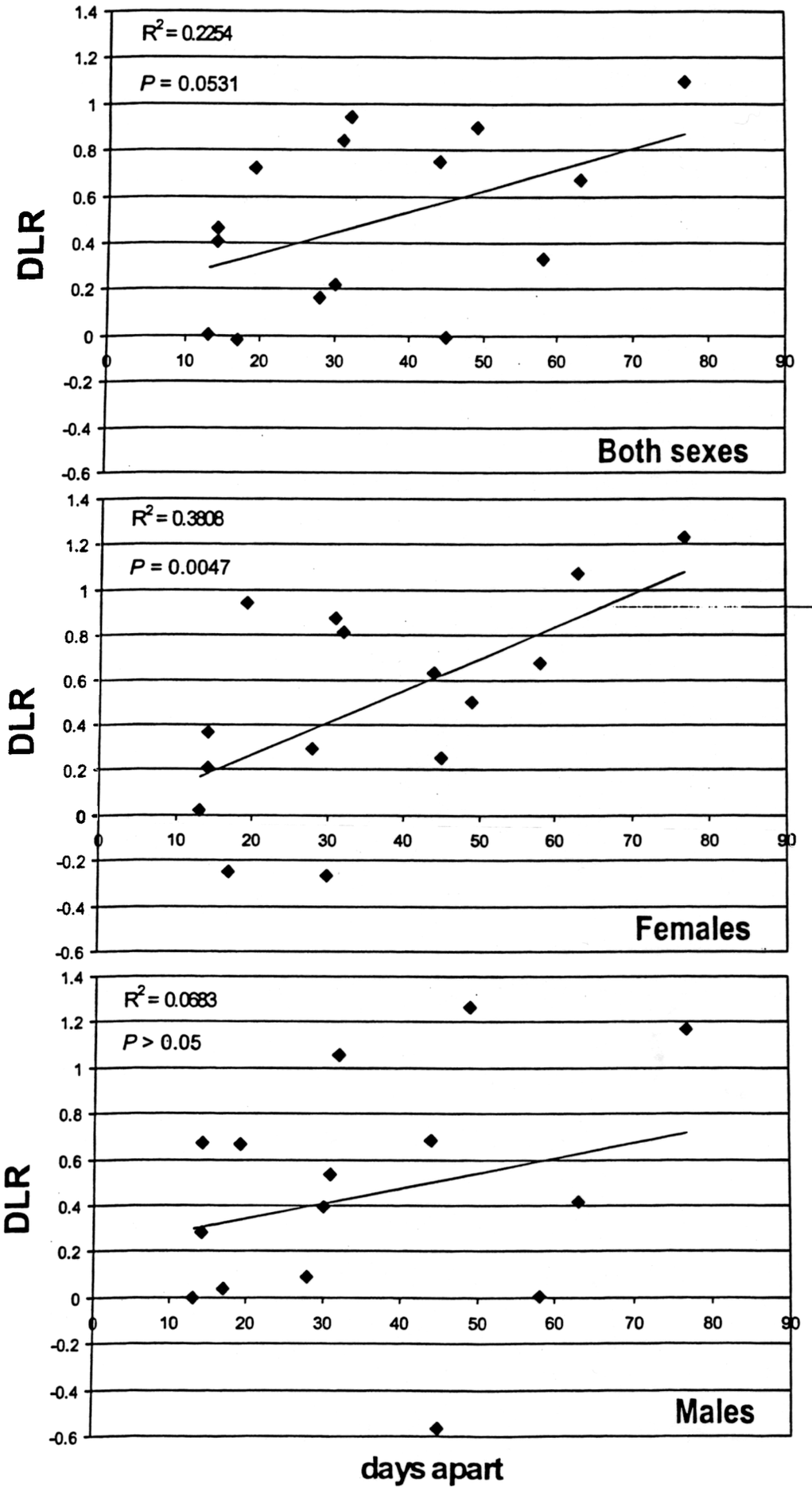

$R^2$ = 0.2254
P = 0.0531
DLR
Both sexes
$R^2$ = 0.3808
P = 0.0047
DLR
Females
$R^2$ = 0.0683
P > 0.05
DLR
Males
days apart

Figure 3. Relationship between genetic distance (DLR) and the number of days separating samples of spawning sockeye from the Cedar R. and its hatchery. Also shown are linear regressions fitted through the data, $R^2$, and the probability that genetic distance and time are not correlated as determined using a Mantel test.

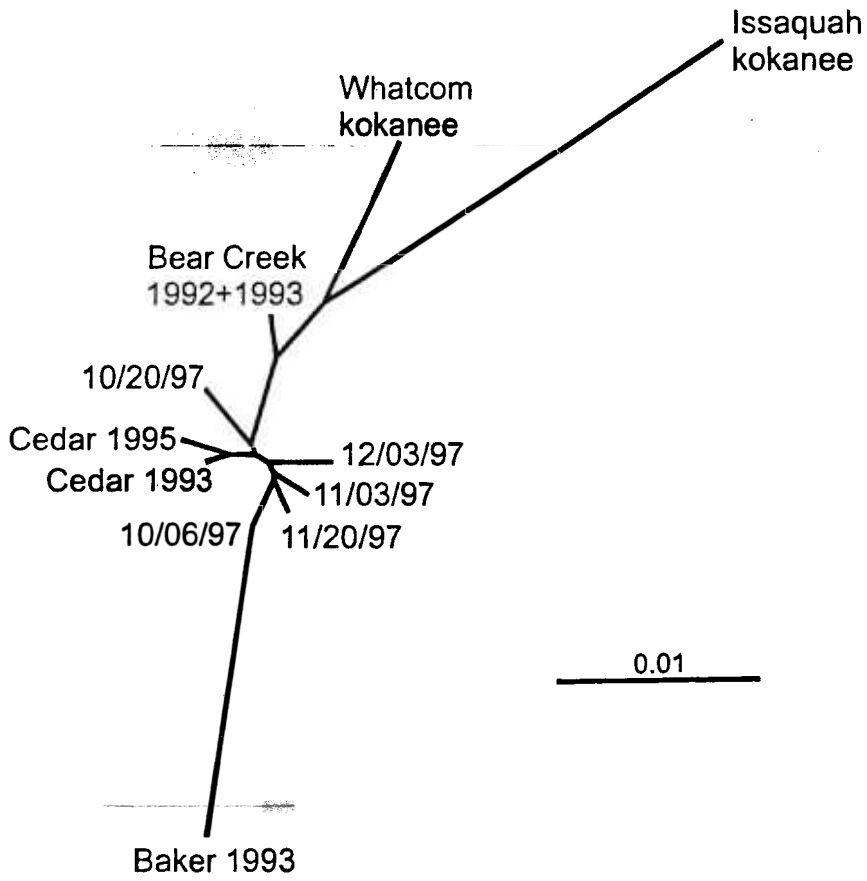



Figure 4. Maximum likelihood tree showing hypothesized relationships of sockeye and kokanee populations based on microsatellite data. Populations indicated only by a date are Cedar R. samples.